

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application. Applicants have submitted a new complete claim set showing any marked up claims with insertions indicated by underlining and deletions indicated by strikeouts and/or double bracketing.

Listing of Claims:

1. (Currently Amended) In a computer data processing system, a method for clustering data in a database comprising:

a) providing a database having a number of data records having both discrete and continuous attributes;

b) grouping together data records in a clustering model from the database which have specified discrete attribute configurations;

c) performing clustering of data records in two phases including a first phase and a second phase, the first phase clustering the data records over a discrete attribute space using an itemset identification, and the second phase clustering continuous attributes using a method for clustering continuous attribute data to produce an intermediate set of data clusters, wherein the first phase precedes the second phase; and

d) merging together clusters from the intermediate set of data clusters to produce a clustering model,

wherein the discrete attributes are Boolean and similarity between configurations is based on a distance between bit patterns of the discrete attributes, and

wherein the step of identifying configurations includes tabulating data records having the same discrete attribute bit pattern and combining the data records from similar configurations before clustering the data records so tabulated based on the continuous attributes.

Type of Response: Amendment
Application Number: 09/886,771
Attorney Docket Number: 163193.01
Filing Date: 21 June 2001

2. (Currently Amended) The method of claim 1 ~~wherein the clustering model includes a table of probabilities for the discrete data attributes of the data records for a cluster and~~ wherein the cluster model for continuous data attributes comprises a mean and a covariance for each cluster.
3. (Original) The method of claim 1 wherein the process of merging of intermediate clusters is ended when a specified number of clusters has been formed.
4. (Original) The method of claim 1 wherein the step of merging of intermediate clusters is ended when a distance between intermediate clusters is greater than a specified minimum distance.
5. (Cancelled)
6. (Currently Amended) The method of claim 1 wherein if one or more of the discrete attributes have more than two possible values ~~and then~~ comprising the step of subdividing a discrete attribute having more than two possible values into multiple Boolean value attributes.
7. (Cancelled)
8. (Currently Amended) In a computer data processing system, a method for clustering data in a database comprising:
- a) providing a database having a number of data records having both discrete and continuous attributes;

Type of Response: Amendment
Application Number: 09/886,771
Attorney Docket Number: 163193.01
Filing Date: 21 June 2001

b) performing a first discrete cluster by itemset identification and identifying a first set of configurations wherein the number of data records of each configuration of said first set of configurations exceeds a threshold number of data records;

c) adding data records from the database not belonging to one of the first set of configurations with a configuration within said first set of configurations to produce a subset of records from the database belonging to configurations in the first set of configurations; and

d) performing a second continuous clustering of the subset of records contained within at least some of the first set of configurations based on the continuous data attributes of records contained within that first set of configurations to produce a clustering model,

wherein the clustering of records from records falling within a configuration of the first set results in a number of intermediate clusters which are merged together to form the cluster model,

wherein intermediate clusters are merged together based on a distance between clusters that is determined based on both continuous and discrete attributes of said intermediate clusters,

wherein the merging of intermediate clusters is performed until a distance between two closest clusters is greater than a threshold distance.

9. (Original) The method of claim 8 wherein the clustering model includes a table of probabilities for the discrete data attributes of the data records for a cluster and wherein the cluster model for continuous data attributes comprises a mean and a covariance for each cluster.

10. (Original) The method of claim 8 wherein an added record not contained within the first set of configurations is added to one of said first set of configurations based on a distance between a smaller configuration to which said added record belongs during

counting of records in different configurations.

11-21. (Cancelled).

22. (Currently Amended) A computer readable medium containing stored instructions for clustering data in a database comprising instructions for:

a) reading records from a database having a number of data records having both discrete and continuous attributes;

b) performing a first clustering of data records from the database which have specified discrete attribute configurations by itemset identification;

c) performing a second clustering of the data records having the same or similar specified discrete attribute configuration based on the continuous attributes to produce an intermediate set of data clusters, wherein the first clustering precedes the second clustering; and

d) merging together clusters from the intermediate set of data clusters to produce a clustering model,

wherein the instructions identify configurations by tabulating data records having the same discrete attribute bit pattern and combining the data records from similar configurations before clustering the data records so tabulated based on the continuous attributes.

23. (Cancelled).

24. (Original) The computer readable medium of claim 22 wherein the instructions end the process of merging of intermediate clusters when a specified number of clusters has been formed.

25. (Original) The computer readable medium of claim 22 wherein the instructions end

the process of merging intermediate clusters when a distance between intermediate clusters is greater than a specified minimum distance.

26. (Original) The computer readable medium of claim 22 wherein the discrete attributes are Boolean and the instructions determine similarity between configurations based on a distance between bit patterns of the discrete attributes.

27-31. (Cancelled).